一种用于交通预测的自适应时空图神经网络

甘 萍[1]，林基明[1,2]，农丽萍[2,3]，王俊义[1]

（1. 桂林电子科技大学 信息与通信学院，广西 桂林 541004；
2. 西安电子科技大学 通信工程学院，西安 710071；
3. 广西师范大学 物理科学与技术学院，广西 桂林 541004）

摘 要：交通预测在城市管理及交通规划中具有重要意义。然而，在交通预测任务中，对复杂动态时空依赖关系的建模仍然具有极大的挑战性。针对以往复杂的神经网络构架在空间维度上所采用的预定义图结构未包含完整交通数据空间信息，且在时间维度上不能很好地捕获交通数据长期依赖关系的问题，提出一种新的时空图神经网络。通过自适应图卷积网络（AGCN）自动捕获节点的特定状态以及自动推断不同节点之间的相互依赖关系，提取更完整的交通数据空间特征，再通过时空长短期记忆网络（ST-LSTM）中的时间记忆模块来提取交通数据的时间特征，捕获短中长期的时间依赖关系。在 PeMSD4 和 PeMSD8 数据集上进行了验证，实验结果表明，所提网络相比基线模型能够更好地提升交通预测性能。

关键词：交通预测；自适应图卷积网络；时空相关性；时空图神经网络；长短期记忆网络

中图分类号：TP391 **文献标志码**：A **文章编号**：1673-808X(2022)05-0094-08

An adaptive spatio-temporal graph neural network for traffic prediction

GAN Ping[1], LIN Jiming[1,2], NONG Liping[2,3], WANG Junyi[1]

(1. School of Information and Communication, Guilin University of Electronic Technology, Guilin 541004, China;
2. School of Telecommunication Engineering, Xidian University, Xi'an 710071, China;
3. College of Physics and Technology, Guangxi Normal University, Guilin 541004, China)

Abstract: Traffic forecasting is of great significance in urban management and traffic planning. However, in the task of traffic prediction, the modeling of complex dynamic spatio-temporal dependence is still a great challenge. For the problem that the neural network can't capture the long-term traffic information in the spatial dimension, the new neural network structure proposed in the past can't capture the complex traffic data in the spatial dimension. Through adaptive graph convolutional network, the specific state of nodes is automatically captured and the interdependence between different nodes is automatically inferred to extract the complete spatial features of traffic data. Then, the time characteristics of traffic data are captured by the time memory module in the spatio-temporal short-term memory network, and the short, medium and long-term time dependence is simulated.

Key words: traffic forecast; adaptive graph convolution network; spatio-temporal correlation; spatio-temporal graph neural network; long-short term memory network

城市化发展对城市交通的流通性和可持续性带来了巨大挑战，有效解决交通问题的智能交通系统（ITS）[1]是目前被广泛关注的研究领域。交通预测作为智能交通系统的一个重要组成部分，不仅为城市交通规划提供科学依据，还在城市交通调度和交通管理中起着至关重要的作用，提高了智能交通系统的决策效率[2]。交通预测任务的实现同时也离不开数据处理、计算机算法、网络优化等技术的支撑。因此，探索高效、实用的交通预测方法不仅能促进多学科之间的交叉应用和发展，而且也成了

收稿日期：2022-03-05
基金项目：国家自然科学基金（61966007）；广西自然科学基金（2020GXNSFAA159105）；广西无线宽带通信与信号处理重点实验室基金（GXKL06190204，GXKL06200116，GXKL06190117）；认知无线电与信息处理教育部重点实验室基金（CRKL180201，CRKL180106）
通信作者：林基明（1970－），男，教授，博士，研究方向为无线通信、移动通信。E-mail：linjm@guet.edu.cn
引文格式：甘萍，林基明，农丽萍，等. 一种用于交通预测的自适应时空图神经网络[J]. 桂林电子科技大学学报，2023，43(2)：94-101.

相关研究领域的热门研究对象，在实际应用中具有非常重要的现实意义[21]。

交通预测是时空预测中的一个典型研究问题，其中交通数据存在着复杂的时空相关性，这使得交通预测任务非常具有挑战性，例如，预测交叉路口未来某时刻的交通状况（交通流量、交通速度和道路占用率等）。通常根据预测时间的长短，交通预测任务可分为短期预测（5～30 min）和中长期预测（30～60 min）[3]。在时间维度上，一条道路此刻的交通状况可能在不久的将来影响该条道路或者其他道路的交通状况。在空间维度上，不同道路之间的交通状况可能会相互影响。因此，研究人员对捕获交通数据时空相关性方面进行了广泛研究。传统方法通过部署时间序列模型进行交通预测，如向量自回归模型（VAR）[4]、自回归集成移动平均模型（ARIMA）[5]、卡尔曼滤波[6]等。然而，这些方法依赖于理想的平稳假设，无法捕捉真实交通动态中大规模交通数据之间的非线性或者复杂的时空相关性。近年来，研究人员转向于研究基于深度学习的方法，通过设计新的神经网络构架来捕捉交通数据的时空相关性，以弥补传统方法的不足，如用于捕获时间依赖性的循环神经网络（RNN）[7]及其变体（长短期记忆网络（LSTM）和门控循环单元(GRU)[8]）或一维卷积神经网络，与用于捕获空间依赖性的卷积神经网络（CNN）[9]相结合来共同捕获交通数据的时空相关性。但这些方式通常适用于欧几里德数据[10]，如图像、网格等，在具有非欧几里德结构的交通网络上仍存在局限性。于是，为了克服上述局限，捕获网络结构特征的图卷积网络（GCN）[11]得到了迅速发展。在图卷积网络中，交通网络被建模成图结构，图中的节点表示交通传感器，边表示传感器之间的连通性，边权值表示传感器之间的欧氏距离。DCRNN[12]、STGCN[13]、ASTGCN[14]、T-GCN[15]等模型通过图神经网络来捕获交通数据的空间依赖性，并通过循环神经网络或卷积神经网络来捕获时间依赖性。这些方法在一定程度上提高了交通预测的精度，但却存在以下问题：1）这些方法采用预定义的图结构[16]，而预定义的图结构未包含完整的空间依赖信息，例如由于交通数据在采集过程中的连接不完整，使部分节点缺少真正的关联；2）网络模型参数空间共享，而不同道路的交通状况不同，共享参数空间使得模型无法很好地捕捉细粒度的特定交通状况（如某段道路封闭或者发生事故）；3）短期预测与长期预测的性能相差较大，不能很好地进行长期预测。AGCRN[17]通过自适应图卷积网络（AGCN）与GRU 相结合的方式来解决空间维度上使用预定义图结构以及共享参数空间所带来的问题，但其在时间维度上并不能很好地处理长序列，即在长期预测方面稍有欠缺。

针对上述问题，设计了一种用于交通预测的自适应时空图神经网络，主要工作概括为利用包含自适应邻接矩阵和自适应节点参数的空间卷积模块提取空间特征。自适应邻接矩阵自动学习节点间的相互依赖关系，捕捉数据中隐藏的空间关系，自适应节点参数根据节点嵌入从所有节点共享的权重和偏差中生成，以捕获特定节点模式；采用时空长短期记忆网络（ST-LSTM）[18]中时间记忆模块来捕获时间特征，增强长序列处理能力；在 2 个公开的 PeMSD4、PeMSD8 交通流量数据集上，与多个基线模型进行对比验证，实验结果表明，本模型能够更好地捕获交通数据时空相关性。。

1　网络模型

1.1　交通预测问题

在交通流预测任务中，通过历史流量数据对未来某时段内的交通流量进行多步预测。交通流量指固定时段内道路某位置通过的车辆数，作为交通数据特征之一，是控制车流和道路规划的重要指标[22]。历史交通数据可表示为交通网络上的时间序列。交通网络可用图的形式表示：$G=(V,E,\mathbf{A})$，V 为图中的节点集合，即路网中的传感器，E 为节点间边的集合，反映了传感器所在道路间的连通性，$\mathbf{A}\in\mathbf{R}^{N\times N}$ 为图 G 的邻接矩阵，当节点 $v_i,v_j\in V$ 且边 $(v_i,v_j)\in E$ 时，邻接矩阵 $\mathbf{A}$ 中的元素 $\mathbf{A}_{ij}=1$，否则为 0。每个时间步长 t，图 G 都有一个表示节点属性的特征矩阵 $\boldsymbol{X}^{(t)}\in\mathbf{R}^{N\times D}$，即图信号。给定图 G 和历史时间步长为 S 的图信号，交通预测问题可表示为通过学习一个映射函数 f，预测未来 T 步长的交通流量，映射关系为

$$\boldsymbol{X}^{(t+1):(t+T)}=f[\boldsymbol{X}^{(t-S):t},G],$$

其中，$\boldsymbol{X}^{(t-S):t}\in\mathbf{R}^{N\times D\times S}$，$\boldsymbol{X}^{(t+1):(t+T)}\in\mathbf{R}^{N\times D\times T}$。

1.2　空间卷积模块

空间卷积模块为自适应图卷积网络，主要包括自适应邻接矩阵和自适应节点参数。其中自适应邻接矩阵通过自动学习节点间的相互依赖关系来捕捉交通数据中隐藏的空间关系；自适应节点参数从所有节点共享的权重池和偏置中生成，以捕获特定节点模式。二者相结合构成的自适应图卷积网络能够获取数据更完整的空间特征。

1.2.1 自适应邻接矩阵

预定义的图结构未包含完整的空间依赖性信息，这可能会导致预测结果出现偏差，而自适应邻接矩阵无需任何先验知识，能从输入数据中自动推断出隐藏的空间依赖关系，借鉴文献[17]中提取空间特征的思路来捕获交通流数据完整的空间相关性。对自适应邻接矩阵部分，首先，为所有节点随机初始化一个可学习的节点嵌入向量 $\boldsymbol{E_A} \in \mathbf{R}^{N\times f_e}$，再通过 $\boldsymbol{E_A}$、$\boldsymbol{E_A^{\mathrm{T}}}$ 来推断每对节点之间空间依赖关系。因此，归一化后的自适应邻接矩阵可表示为

$$\widetilde{\boldsymbol{A}} = \boldsymbol{D}^{-\frac{1}{2}}\boldsymbol{A}\boldsymbol{D}^{-\frac{1}{2}} = \mathrm{Softmax}(\mathrm{ReLU}(\boldsymbol{E_A} \cdot \boldsymbol{E_A^{\mathrm{T}}})),$$

其中：$\boldsymbol{E_A}$ 的每行表示一个节点嵌入；N 为节点数；f_e 为节点嵌入的维度；度矩阵 $\boldsymbol{D} \in \mathbf{R}^{N\times N}$ 为对角矩阵；$\boldsymbol{A} \in \mathbf{R}^{N\times N}$ 为图的邻接矩阵；ReLU 为激活函数，Softmax 函数对自适应邻接矩阵进行归一化。在模型训练过程中，$\boldsymbol{E_A}$ 通过自动更新来学习不同流量序列之间的隐藏依赖关系，同时获得图卷积的邻接关系。

1.2.2 自适应节点参数

在交通网络中，仅捕获节点间的共享模式对于交通预测是不够准确的。因为路网结构中 2 个相邻节点的特征可能由于其特定属性，如天气、事故等，而在某个特定时间段呈现不同的交通情况，或者 2 个不相邻节点的特征呈现相反的情况。因此，有必要为每个节点设定独立的参数空间来学习特定节点模式。一般的图卷积网络（GCN）[9] 运算方式为 $\boldsymbol{Z} = (\boldsymbol{I}_N + \boldsymbol{D}^{-1/2}\boldsymbol{A}\boldsymbol{D}^{-1/2})\boldsymbol{X}\boldsymbol{\Theta} + \boldsymbol{b}$，其中：$\boldsymbol{X} \in \mathbf{R}^{N\times C}$ 为输入；C 为特征数；$\boldsymbol{Z} \in \mathbf{R}^{N\times F}$ 为输出；F 为输出维度；$\boldsymbol{\Theta} \in \mathbf{R}^{C\times F}$ 和 $\boldsymbol{b} \in \mathbf{R}^{F}$ 分别为可学习的权重和偏差。对自适应节点参数部分，分别定义一个节点嵌入矩阵 $\boldsymbol{E}_G \in \mathbf{R}^{N\times f}$ 和权重池 $\boldsymbol{W}_G \in \mathbf{R}^{f\times C\times F}$，$f$ 为嵌入维度，且 $f \ll N$，$\boldsymbol{\Theta} = \boldsymbol{E}_G \cdot \boldsymbol{W}_G$。该过程可看作从所有流量序列的一组候选模式中学习节点特定模式，即节点嵌入 $\boldsymbol{E}_G^i$ 从大型共享权重池 $\boldsymbol{W}_G$ 中获取节点 i 的参数 $\boldsymbol{\Theta}^i$，$\boldsymbol{b}$ 的操作相同。采用自适应节点参数的 GCN 公式：

$$\boldsymbol{Z} = (\boldsymbol{I}_N + \boldsymbol{D}^{-\frac{1}{2}}\boldsymbol{A}\boldsymbol{D}^{-\frac{1}{2}})\boldsymbol{X}\boldsymbol{E}_G\boldsymbol{W}_G + \boldsymbol{E}_G\boldsymbol{b}_G。$$

1.3 时间卷积模块

时间特征的提取采用时空长短期记忆网络（ST-LSTM）[18] 的时间记忆模块。通过卷积层取代 LSTM 中各个门的线性操作，类似于 ConvLSTM，这种时间特征提取方式能更好地处理非线性的时空序列数据，解决梯度消失/爆炸问题，增强长序列处理能力，提高模型长期预测性能。

$$g_t = \tanh(\boldsymbol{W}_{xg} * \boldsymbol{X}_t + \boldsymbol{W}_{hg} * \boldsymbol{H}_{t-1}^l + \boldsymbol{b}_g),$$
$$i_t = \sigma(\boldsymbol{W}_{xi} * \boldsymbol{X}_t + \boldsymbol{W}_{ht} * \boldsymbol{H}_{t-1}^l + \boldsymbol{b}_i),$$
$$f_t = \sigma(\boldsymbol{W}_{xf} * \boldsymbol{X}_t + \boldsymbol{W}_{hf} * \boldsymbol{H}_{t-1}^l + \boldsymbol{b}_f),$$
$$\boldsymbol{C}_t^l = f_t \odot \boldsymbol{C}_{t-1}^l + i_t \odot g_t,$$
$$o_t = \sigma(\boldsymbol{W}_o * [\boldsymbol{X}_t;\boldsymbol{H}_{t-1}^l;\boldsymbol{C}_t^l;\Delta]),$$
$$\boldsymbol{H}_t^l = o_t \odot \tanh(\boldsymbol{W}_{1\times 1} * [\boldsymbol{C}_t^l;\Delta])。$$

其中：$\boldsymbol{C}_t^l$ 为 t 时刻 l 层的时间信息记忆单元；g_t、i_t 为输入门，选择性记录新信息到记忆单元；f_t 为遗忘门，根据数据特征的重要性选择前一时刻记忆单元 C_{t-1}^l 中的关键信息保留至当前记忆单元 $\boldsymbol{C}_t^l$；o_t 为输出门，输出所需要的信息；$\boldsymbol{H}_t^l$ 为 t 时刻 l 层的输出信息；$*$ 为卷积操作；$\odot$ 为哈达玛积；Δ 为自适应图卷积捕获的空间特征信息。

1.4 模型框架

自适应图卷积长短期记忆网络模型（AGC-LSTM）如图 1 所示，AGCN 模块进行空间特征的提取，$\boldsymbol{M}_t^l$ 为 t 时刻 l 层的空间信息记忆单元。将时间记忆和空间记忆使用共享的输出门 o_t，实现了无缝记忆融合，可有效地对时空序列数据的变化进行建模。整个网络模型可表示为

$$g_t = \tanh(\boldsymbol{W}_{xg} * \boldsymbol{X}_t + \boldsymbol{W}_{hg} * \boldsymbol{H}_{t-1}^l + \boldsymbol{b}_g),$$
$$i_t = \sigma(\boldsymbol{W}_{xi} * \boldsymbol{X}_t + \boldsymbol{W}_{ht} * \boldsymbol{H}_{t-1}^l + \boldsymbol{b}_i),$$
$$f_t = \sigma(\boldsymbol{W}_{xf} * \boldsymbol{X}_t + \boldsymbol{W}_{hf} * \boldsymbol{H}_{t-1}^l + \boldsymbol{b}_f),$$
$$\boldsymbol{C}_t^l = f_t \odot \boldsymbol{C}_{t-1}^l + i_t \odot g_t,$$
$$\widetilde{\boldsymbol{A}} = \mathrm{Softmax}(\mathrm{ReLU}(\boldsymbol{E_A} \cdot \boldsymbol{E_A^{\mathrm{T}}})),$$
$$\boldsymbol{M}_t^l = \widetilde{\boldsymbol{A}}[\boldsymbol{X};\boldsymbol{M}_t^{l-1}]\boldsymbol{E}\boldsymbol{W}_{xm} + \boldsymbol{E}\boldsymbol{b}_{xm},$$
$$o_t = \sigma(\boldsymbol{W}_o * [\boldsymbol{X}_t;\boldsymbol{H}_{t-1}^l;\boldsymbol{C}_t^l;\boldsymbol{M}_t^l]),$$
$$\boldsymbol{H}_t^l = o_t \odot \tanh(\boldsymbol{W}_{1\times 1} * [\boldsymbol{C}_t^l;\boldsymbol{M}_t^l])。$$

通过堆叠多层 AGC-LSTM，捕获特定节点的时空模式，实现多步流量预测，即输入历史数据 $\boldsymbol{H} \in \mathbf{R}^{N\times m}$，通过网络模型将 $\mathbf{R}^{N\times m}$ 映射到 $\mathbf{R}^{N\times \tau}$ 来直接获得所有节点未来 τ 步长的交通流量。采用 L1 损失函数作为模型训练目标，

$$L(\boldsymbol{W}_\theta) = \sum_{i=t+1}^{i=t+\tau} |\boldsymbol{X}_{:,i} - \boldsymbol{Y}_{:,i}|,$$

其中：$\boldsymbol{W}_\theta$ 为网络中所有可训练的参数；$\boldsymbol{X}_{:,i} \in \mathbf{R}^{N\times 1}$ 为第 i 个时间步的真实值；$\boldsymbol{Y}_{:,i} \in \mathbf{R}^{N\times 1}$ 为第 i 个样本的预测值。

2 实验探究

2.1 实验设置

为了验证所提模型的优越性，在 PeMSD4[14] 和

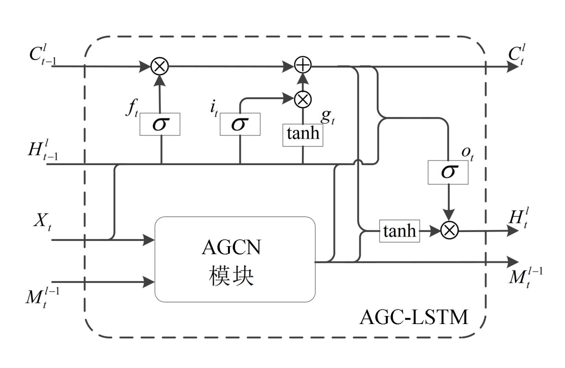


图 1　网络模型 AGC-LSTM

PeMSD8[17]数据集上进行实验验证。PeMSD4 数据集收集了 2018 年 1 月 1 日至 2018 年 2 月 28 日旧金山湾区 307 个交通监测器所监测的交通流量信息，PeMSD8 数据集收集了 2016 年 7 月 1 日至 2016 年 8 月 31 日圣贝纳迪诺地区 170 个环路探测器收集的交通流量信息，交通流量数据每 5 min（即时间步长为 1）汇聚一次。具体数据集统计信息如表 1 所示。

数据集中缺失的值采用线性插值填充，并采用标准归一化方式对数据集进行归一化处理，使训练过程更加稳定。在 2 个数据集上均使用 12 步长（历史时间步长为 12）的历史数据，预测未来 12 步长的交通流量信息。数据集划分为 6∶2∶2的训练集、验证集和测试集。采用 Adam 优化器进行优化，Adam 优化器计算高效，对内存需求少，很适合应用于大规模数据及参数的场景。迭代次数为 100，bitch size 大小为 64，初始学习率为 0.001，并使用 patience 为 15 的早停方法。

表 1　数据集统计信息

数据集	# Nodes	# Edges	# Time Steps
PeMSD4	307	340	16 992
PeMSD8	107	277	17 856

2.2　评估指标和基线

采用 3 个被广泛使用于交通预测任务中的评估指标来衡量不同对比模型的性能，分别为平均绝对误差（MAE）、均方根误差（RMSE）和平均绝对百分比误差（MAPE），

$$\mathrm{MAE}=\frac{1}{N}\sum_{i=1}^{N}|\boldsymbol{X}_{:,i}-\boldsymbol{Y}_{:,i}|, \tag{19}$$

$$\mathrm{RMSE}=\sqrt{\frac{1}{N}\sum_{i=1}^{N}(\boldsymbol{X}_{:,i}-\boldsymbol{Y}_{:,i})^2}, \tag{20}$$

$$\mathrm{MAPE}=\frac{100\%}{N}\sum_{i=1}^{N}\left|\frac{\boldsymbol{X}_{:,i}-\boldsymbol{Y}_{:,i}}{\boldsymbol{X}_{:,i}}\right|。 \tag{21}$$

MAE、RMSE 和 MAPE 指标用于测量预测误差，误差越小，表明预测性能越好。

2.3　实验结果

不同方法在 12 个步长预测区间内的总体性能比较结果如表 2 所示。从表 2 可看出，模型 AGC-LSTM 在 PeMSD4 和 PeMSD8 两个数据集上均获得了良好效果。与 VAR 和 SVR 等传统机器学习方法相比，AGC-LSTM 的预测性能有了大幅度提升，其在 2 个数据集上的表现均较传统方法更好，说明使用时空图神经网络的模型能更好地处理时空序列数据。与 LSTM 方法相比，AGC-LSTM 同样具有更好的预测性能，且性能提升较大，说明相较于一般的卷积神经网络，AGC-LSTM 捕获数据时空相关性的能力更强。在同样使用时空图神经网络的方法中，AGC-LSTM 与 ASTGCN 相比，在 PeMSD4 数据集上，MAE 降低了 3.9%，在 PeMSD8 数据集上，MAE 降低了 1.75%；与 STSGCN 相比，在 PeMSD4 数据集上，MAE 降低了 1.24%，在 PeMSD8 数据集上，MAE 降低了 1.07%。这充分说明与使用预定义图结构的模型相比，AGC-LSTM 能更好地捕获交通数据的时空相关性。与同样使用自适应图卷积网络（AGCN）的 AGRCN 相比，AGC-LSTM 的 3 个误差值都更小，说明其捕获交通数据时间依赖性的能力更强。

表 2　PeMSD4 和 PeMSD8 数据集上不同方法在 12 步长预测区间内的平均性能比较

模型	PeMSD4			PeMSD8		
	MAE	RMSE	MAPE/%	MAE	RMSE	MAPE/%
VAR	24.54	38.61	17.24	19.19	29.81	13.10
SVR	28.70	44.56	19.20	28.70	36.16	14.64
LSTM	27.14	41.59	18.20	23.25	34.06	14.20
DCRNN	21.22	33.44	14.17	16.82	26.36	10.92
STGCN	21.16	34.89	13.83	17.50	27.09	11.29
ASTGCN	22.93	35.22	16.56	18.25	28.06	11.64
STSGCN	21.19	33.65	13.90	17.13	26.86	10.96
AGCRN	19.83	32.26	12.97	15.95	25.22	10.09
AGC-LSTM	19.30	31.47	12.66	15.56	25.02	9.89

在 PeMSD4 和 PeMSD8 数据集上，12 个预测步长区间内各步长的预测性能如图 2～7 所示。从图 2～7 可看出，模型 AGC-LSTM 在 12 个预测步长区间，除第 1 步长外，在其他步长均具有良好的预测性能，且在 2 个数据集上均表现良好，能够更好地平衡短期和中长期预测，即在保证短期流量预测性能的同时，能够提升一定程度的长期预测性能，说明 AGC-

LSTM 的多步预测能力更强，具有较好的泛化力。

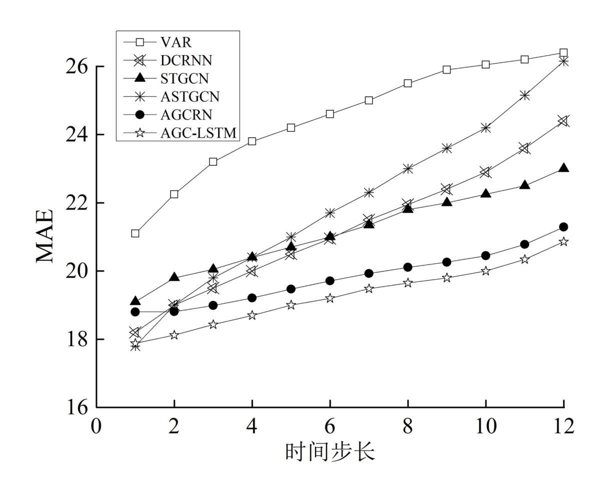


图 2 PeMSD4 数据集上不同方法的 MAE 值

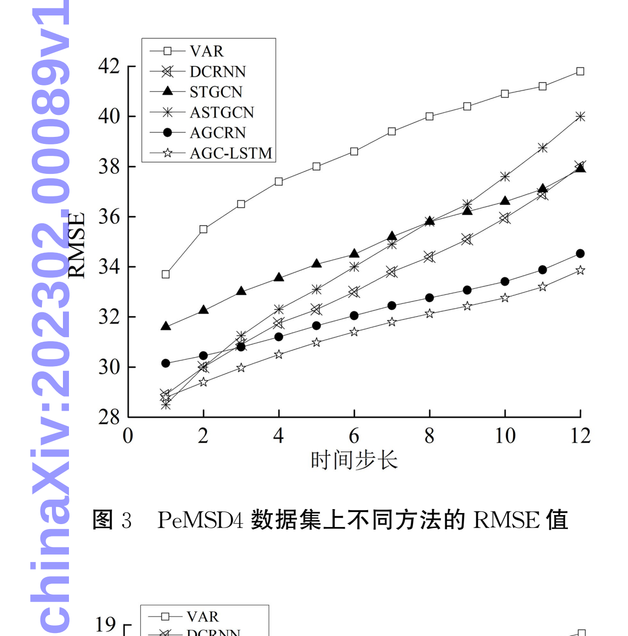


图 3 PeMSD4 数据集上不同方法的 RMSE 值

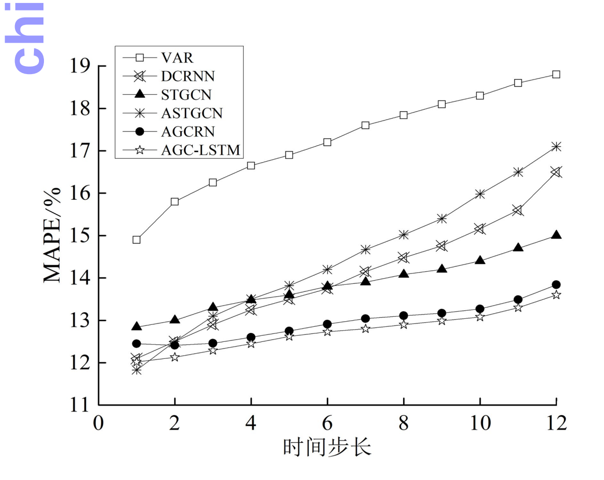


图 4 PeMSD4 数据集上不同方法的 MAPE 值

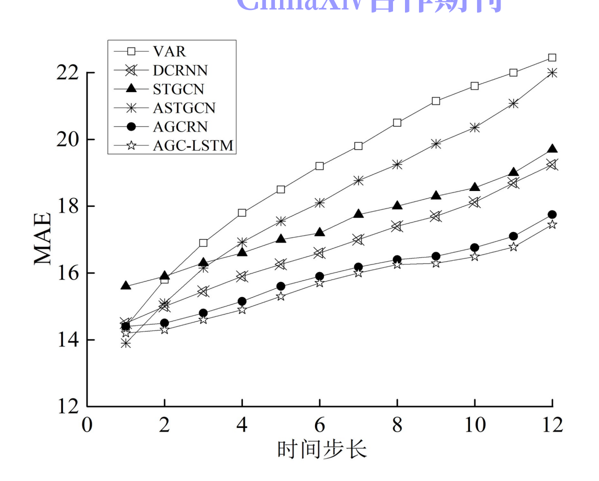


图 5 PeMSD8 数据集上不同方法的 MAE 值

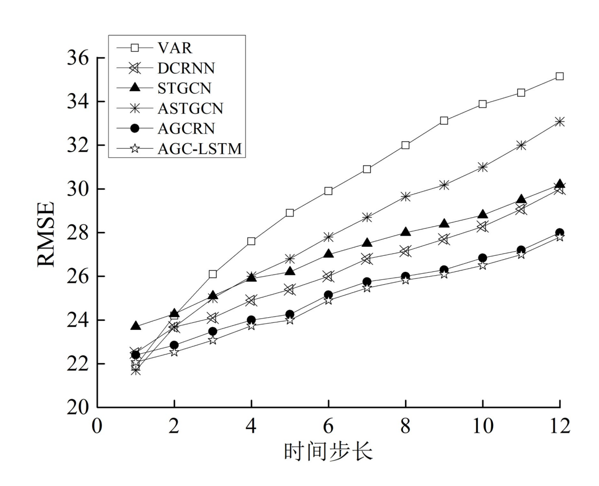


图 6 PeMSD8 数据集上不同方法的 RMSE 值

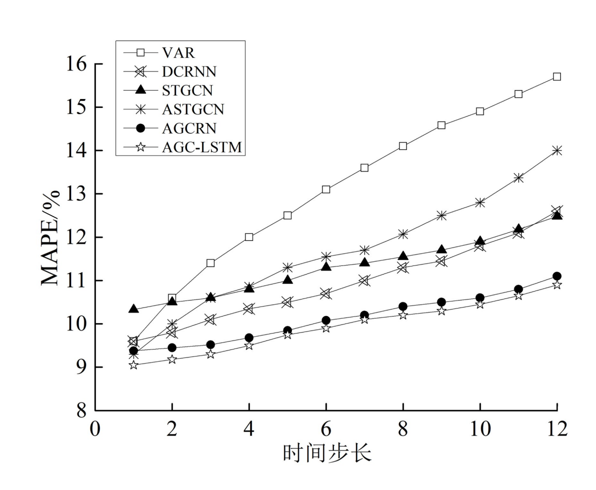


图 7 PeMSD8 数据集上不同方法的 MAPE 值

2.4 嵌入维度分析

在 AGC-LSTM 中，自适应图卷积的节点嵌入维度是一个重要参数，其大小直接影响模型提取时空信息的能力。以 PeMSD4 数据集为测试对象，考虑到计算成本，选择网络模型层数 K 分别为 1、2、3，且在不同 K 值下，将嵌入维度分别设为 4、6、9、12、15，AGC-LSTM 在 12 步长预测区间内的平均 MAE 值和平均 MAPE 值(以下实验都是取各评估指标在 12

预测步长区间内的平均值）如图 8 所示。从图 8 可看出，在不同 K 值下，嵌入维度不同，模型的性能也不同，嵌入维度过大或者过小都会影响模型的性能，在考虑到嵌入维度越大包含数据信息越多的同时，也要考虑到模型参数的增加，从而加大模型优化的难度。综合 K 值来看，当嵌入维度为 12 时，AGC-LSTM 能取得最佳性能。

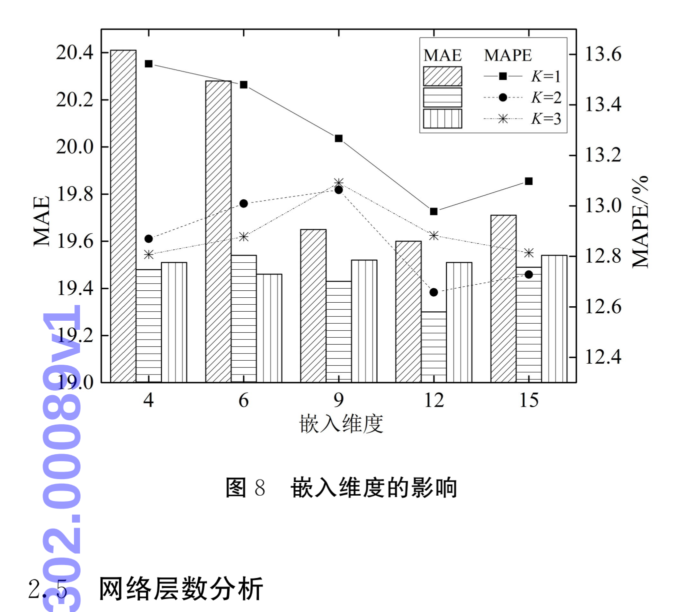


图 8　嵌入维度的影响

2.5　网络层数分析

AGC-LSTM 堆叠层数对捕获交通数据的时空相关性的能力有一定的影响，但随着网络层数的增加，模型参数、计算成本和计算复杂度也会增加，以 PeMSD4 数据集为测试对象，考虑以上问题，选择分析嵌入维度为 12 时的 1、2、3 层 AGC-LSTM，如图 9 所示。从图 9 可看出，模型堆叠层数不同，预测性能不同，当网络层数为 2 时，MAE 值和 MAPE 值最小，此时 AGC-LSTM 提取时空相关性的能力相对最强，具有最佳预测性能。

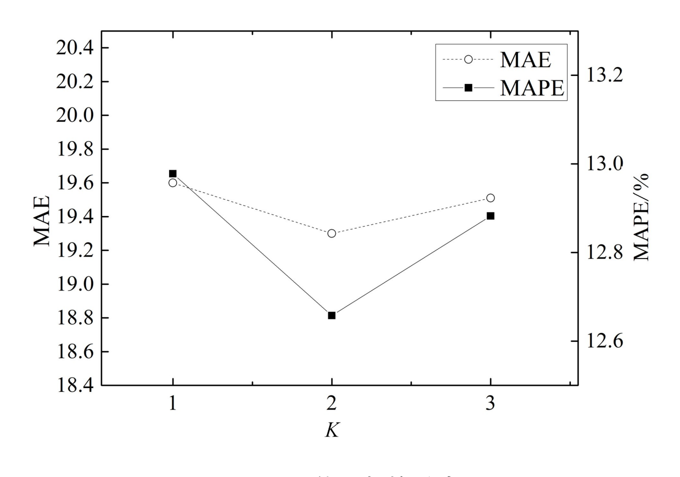


图 9　网络层数的影响

2.6　网络层数分析

在 PeMSD4 数据集上，进一步评估了 AGC-LSTM 中自适应图卷积模块的性能。使用一般的 GCN 来提取空间特征，并以相同方式与时间卷积模块相结合为 GC-LSTM，以共同提取时空特征，再与本方法进行对比，对比结果如图 10 所示。从图 10 可看出，AGC-LSTM 中的自适应图卷积模块能够更好地学习特定节点模式，可从历史观测数据中获取足够的空间特征信息，因此，其预测性能优于 GC-LSTM。

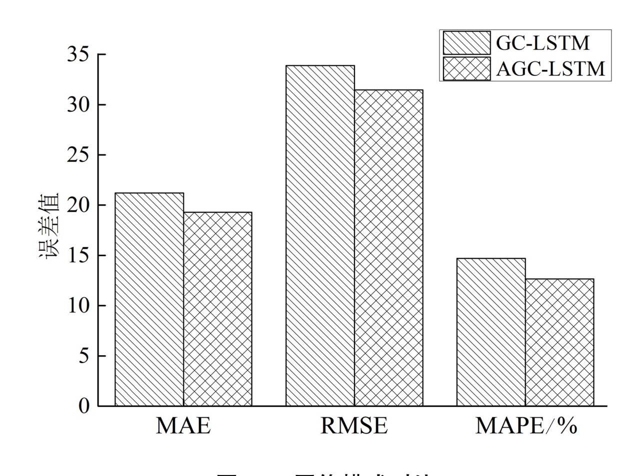


图 10　网络模式对比

2.7　计算复杂度分析

在 PeMSD4 数据集上，AGC-LSTM 与 DCRNN、STGCN、ASTGCN、AGCRN 的参数量和训练时间如表 3 所示。从表 3 可看出，AGC-LSTM 的参数量高于 DCRNN 和 STGCN，这是因为为了更好地学习特定节点模式，增加了节点嵌入维度，并设置为 12。ASTGCN 中添加了时空注意力机制，以获取更精确的时空特征，但也增加了参数量和训练时间，AGCRN 采用自适应卷积递归算法，增加了模型参数量。当 AGC-LSTM 层数为 2 时，模型参数较 ASTGCN 和 AGCRN 要少，且预测性能也较好，其训练时间略长，这是因为带有 3 个卷积门结构的时间卷积模块经过堆叠，在增强时间信息提取性能的同时，增加了计算复杂度。当 AGC-LSTM 层数为 1 时，模型参数较 DCRNN、STGCN、ASTGCN、AGCRN 更少，训练时间与 RMSE 值也较小。综合考虑参数量和预测精度，2 层 AGC-LSTM 的预测性能更佳，计算成本适中。

表 3 PeMSD4 数据集的计算复杂度

模型	#Parameters	#Training Time epoch/s	RMSE
DCRNN	149 057	46.47	33.44
STGCN	211 596	21.56	34.89
ASTGCN	450 031	54.49	35.22
AGCRN	748 810	40.67	32.26
AGC-LSTM (layers=1)	138 096	28.60	31.82
AGC-LSTM (layers=2)	384 624	57.81	31.47

3 结束语

自适应图卷积长短期记忆网络 AGC-LSTM 在复杂的交通预测中取得了良好的效果。AGC-LSTM 由自适应图卷模块和时间卷积模块两部分组成，自适应图卷积模块能够自动捕获节点的特定状态并自动推断不同节点之间的相互依赖关系，时间卷积模块能较好地提取短期、中长期的时间信息，解决梯度消失/爆炸的问题。在 PeMSD4 和 PeMSD8 数据集上证明了本方法对交通预测任务的有效性。下一步工作将优化网络，降低复杂度。

参考文献：

[1] MU S D, XIONG Z Y, TIAN Y X. Intelligent traffic control system based on cloud computing and big data mining[J]. IEEE Transactions on Industrial Informatics, 2019, 15(12): 6583-6592.

[2] YANG H F, DILLON T S, CHANG E, et al. Optimized configuration of exponential smoothing and extreme learning machine for traffic flow forecasting[J]. IEEE Transactions on Industrial Informatics, 2018, 15(1): 23-34.

[3] VLAHOGIANNI E I, KARLAFTIS M G, GOLIAS J C. Optimized and meta-optimized neural networks for short-term traffic flow prediction: a genetic approach[J]. Transportation Research Part C: Emerging Technologies, 2005, 13(3): 211-234.

[4] MALEKI M, WRAITH D E, MAHMOUDI M R, et al. Asymmetric heavy-tailed vector auto-regressive processes with application to financial data[J]. Journal of Statistical Computation and Simulation, 2020, 90(2): 324-340.

[5] LIU W, ZHENG Y, CHAWLA S, et al. Discovering spatio-temporal causal interactions in traffic data streams[C]//Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York, NY: ACM Press, 2011: 1010-1018.

[6] KUMAR S V. Traffic flow prediction using Kalman filtering technique[J]. Procedia Engineering, 2017, 187: 582-587.

[7] MA Xiadei, DAI Zhuang, HE Zhengbing, et al. Learning traffic as images: a deep convolutional neural network for large-scale transportation network speed prediction[J]. Sensors, 2017, 17(4): 818.

[8] FU R, ZHANG Z, LI L. Using LSTM and GRU neural network methods for traffic flow prediction[C]//2016 31st Youth Academic Annual Conference of Chinese Association of Automation. Piscataway, NJ: IEEE Press, 2016: 324-328.

[9] ZHOU D X. Theory of deep convolutional neural networks: downsampling[J]. Neural Networks, 2020, 124: 319-327.

[10] LI Chensheng, QIN Xiaowei, XU Xiaodong, et al. Scalable graph convolutional networks with fast localized spectral filter for directed graphs[J]. IEEE Access, 2020, 8: 105634-105644.

[11] LIU Yanbei, WANG Qi, WANG Xiao, et al. Community enhanced graph convolutional networks[J]. Pattern Recognition Letters, 2020, 138: 462-468.

[12] LI Yaguang, YU Rose, SHAHABI C, et al. Diffusion convolutional recurrent neural network: data-driven traffic forecasting[C]//2018 International Conference on Learning Representations. [S. l.]: ICLR Press, 2018: 1-16.

[13] YU Bing, YIN Haoeng, ZHU Zhanxing. Spatio-temporal graph convolutional networks: a deep learning framework for traffic forecasting[C]//Proceedings of the 27th International Joint Conference on Artificial Intelligence. New York, NY: ACM, 2018: 3634-3640.

[14] GUO S, LIN Y, FENG N, et al. Attention based spatial-temporal graph convolutional networks for traffic flow forecasting[C]//Proceedings of the AAAI Conference on Artificial Intelligence. Menlo Park: AAAI, 2019, 33(1): 922-929.

[15] ZHAO Ling, SONG Yujiao, ZHANG Chao, et al. T-GCN: a temporal graph convolutional network for traffic prediction[J]. IEEE Transactions on Intelligent Transportation Systems, 2020, 21(9): 3848-3858.

[16] SEO Y, DEFFERRARD M, VANDERGHEYNST P, et al. Structured sequence modeling with graph convolutional recurrent networks[C]//International Conference on Neural Information Processing. Berlin: Springer, 2018: 362-373.

[17] BAI Lei, YAO Lina, LI Can, et al. Adaptive graph con-

volutional recurrent network for traffic forecasting[J]. Advances in Neural Information Processing Systems, 2020, 33: 17804-17815.

[18] WANG Yunbo, LONG Mingsheng, WANG Jianmin, et al. PredRNN: recurrent neural networks for predictive learning using spatiotemporal LSTMs[C]//Proceedings of the 31st Conference on Neural Information Processing Systems. New York, NY: ACM, 2017: 879-888.

[19] KAVITHA S, VARUNA S, RAMYA R. A comparative analysis on linear regression and support vector regression[C]//2016 Online International Conference on Green Engineering and Technologies. Piscataway, NJ: IEEE Press, 2016: 1-5.

[20] SONG Chao, LIN Youfang, GUO Shengnan, et al. Spatial-temporal synchronous graph convolutional networks: a new framework for spatial-temporal network data forecasting[C]//Proceedings of the AAAI Conference on Artificial Intelligence. Palo Alto: AAAI, 2020: 914-921.

[21] YAO Haipeng, GAO Pengcheng, WANG Jingjing, et al. Capsule network assisted IoT traffic classification mechanism for smart cities[J]. IEEE Internet of Things Journal, 2019, 6(5): 7515-7525.

[22] MAERIVOET S, DE MOOR B. Traffic flow theory [J]. Physics, 2005, 1(1/2): 5-7.

编辑：张所滨